ZooKeys 1264: 73–93 (2025)
DOI: 10.3897/zookeys.1264.134756

Research Article

A digitization workflow of dry-pinned collections of Lepidoptera

Laurel Kaminsky[1], Stacey Huber[2], Trudi Deuel[2], Ngoc-Nhu V. Tran[2], Erin Jane Howard[2], Aaron Leopold[2], Anupama Priyadarshini[2], David Plotkin[2], Akito Y. Kawahara[2]

1 *Marston Science Library, Smathers Libraries, University of Florida, 444 Newell Drive, Gainesville, Florida 32608, USA*

2 *McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, 3215 Hull Road, Gainesville, Florida 32611, USA*

Corresponding author: Laurel Kaminsky (kaminsky@ufl.edu)

Abstract

Butterflies and moths (Lepidoptera) are one of the most commonly found insect groups in museum collections. Yet many specimens still lack digital data and few digitization workflows are available. Here we present a digitization workflow for natural history collections that can be widely applicable for any museum or dried pinned-specimen collection of Lepidoptera. Our workflow consists of pre-imaging preparation, usage of a copy stand for imaging pinned specimens and data labels, image processing, and transcription, the latter two which utilize Python scripts for optimization.

Key words: Butterfly, copy stand, digitization, imaging, moth, museum

OPEN ACCESS

Academic editor: Shinichi Nakahara
Received: 3 March 2025
Accepted: 4 November 2025
Published: 15 December 2025

ZooBank: https://zoobank.org/B5886647-DF66-488C-9182-A83798E09203

Citation: Kaminsky L, Huber S, Deuel T, Tran N-NV, Howard EJ, Leopold A, Priyadarshini A, Plotkin D, Kawahara AY (2025) A digitization workflow of dry-pinned collections of Lepidoptera. ZooKeys 1264: 73–93. https://doi.org/10.3897/zookeys.1264.134756

Copyright: © Laurel Kaminsky et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Introduction

Arthropods are a diverse group of organisms with morphologically distinctive life stages (Giribet and Edgecomb 2012). There are an estimated 300 million adult arthropod specimens across 223 North American collections (Cobb et al. 2019). These specimens and their data are needed to answer questions in many fields of biology, ranging from the genetic level to the study of entire ecosystems (Short et al. 2018). Currently, the majority of data is inaccessible to researchers around the world because many collections are not digitized. Digitization creates a record of collections that enables global and widespread access to data, broadens the scale and scope of collections-based research, and enhances accessibility to natural history collections (NHCs) data (Nelson and Ellis 2018; Hedrick et al. 2020).

Methods that NHCs use for record keeping have changed over the years. Recording label data has historically focused on tracking specimens in a collection. Modern digitization approaches are focused on capturing and sharing label data, metadata, and research grade high resolution images of specimens in NHCs with researchers around the world through data aggregators such as Integrated Digitized Biocollections (iDigBio 2024) and Global Biodiversity Information Facility (GBIF 2024).

Due to these advancements, exponentially more data are being gathered from specimens to answer a diverse array of questions (Ball-Damerow et al. 2019). Specimen image and label data can be used to assemble species lists

for specific regions (Terry et al. 2023), provide insight into biological phenomena such as biogeography (Bedford et al. 2012; Earl et al. 2021), phenology (Wells and Tonkyn 2014; Echevarría Ramos and Hulshof 2019; Li et al. 2019; Belitz et al. 2023), fungal pathogens (May and Ristaino 2004), climate change (Kharouba et al. 2019), morphology (Holmes et al. 2016; Owens et al. 2020; Hamilton et al. 2022), and behavior (Aiello et al. 2021; Childers et al. 2023). Images can be used to train neural networks to identify arthropods, as has been demonstrated in analyses of Hymenoptera (Buschbacher et al. 2020), Hemiptera (Popkov et al. 2022), and Lepidoptera (Almryad and Kutucu 2020; Cunha et al. 2023; Qin et al. 2024).

Digitization workflows usually consist of five steps: pre-imaging preparation, imaging, image processing, transcription, and georeferencing (Nelson et al. 2012). The steps of the digitization workflow depend on the organism method of preservation, the research question, and the NHC's institutional protocols. Key features of most workflows are the use of a light source and a camera to image the specimen and/or data labels (Fig. 1). The copy stand is a widely used tool consisting of a flat platform with a central mount for adjusting camera height and integrated or separate lighting. It is particularly suited for photographing specimens preserved as flat sheets or packets, such as herbarium and fungarium specimens, as well as pinned arthropods (Nelson et al. 2015; Thiers et al. 2016; Harris and Marsico 2017; Takano et al. 2019).

In NHCs, the majority of arthropod dry collections remain undigitized (Cobb et al. 2019). Digitizing these collections and providing access to these data is challenging because of the sheer volume of specimens, diversity of life stages, and methods (e.g. pinned, papered, fluid) of preservation and storage. In addition, a pinned-arthropod specimen often obscures label data located beneath it on the pin. Arthropod digitization can be perceived as still being in its infancy but rapidly evolving. Digital records of specimens in arthropod collections typically consist of images of the specimens and transcriptions of associated label data.

For arthropods, there are few published peer-reviewed digitization workflows that use copy stands or scanners. One Lepidoptera workflow is the Natural History Museum's iCollections project (Paterson et al. 2016; Blagoderov et al. 2017), which covers specimen selection, imaging, transcription, georeferencing, and data management. A copy stand variant workflow was used to digitize papered swallowtail butterflies (Caspers et al. 2019). Two fluid digitization papers on arthropods have been published. One was designed for stonefly digitization and entails imaging a specimen in a petri dish and associated data labels underneath (DeWalt et al. 2018). The second workflow utilized a flatbed scanner and 3D-printed boxes to image specimens and data labels for caddisflies in fluid (Mendez et al. 2018).

The vast number of specimens in many NHCs makes complete digitization challenging, especially for large and rapidly growing collections, when using commonly employed tools like a copy stand (Cobb et al. 2019). Some museums are now employing fast, automated methods that include conveyor belts (Tegelberg et al. 2014, 2017) which allow simultaneous preparation and imaging of specimens. Further automation of the copy stand digitization process, including imaging, transcription, and georeferencing, has been proposed for dry collections through the National Science Foundation (NSF)-funded Light-

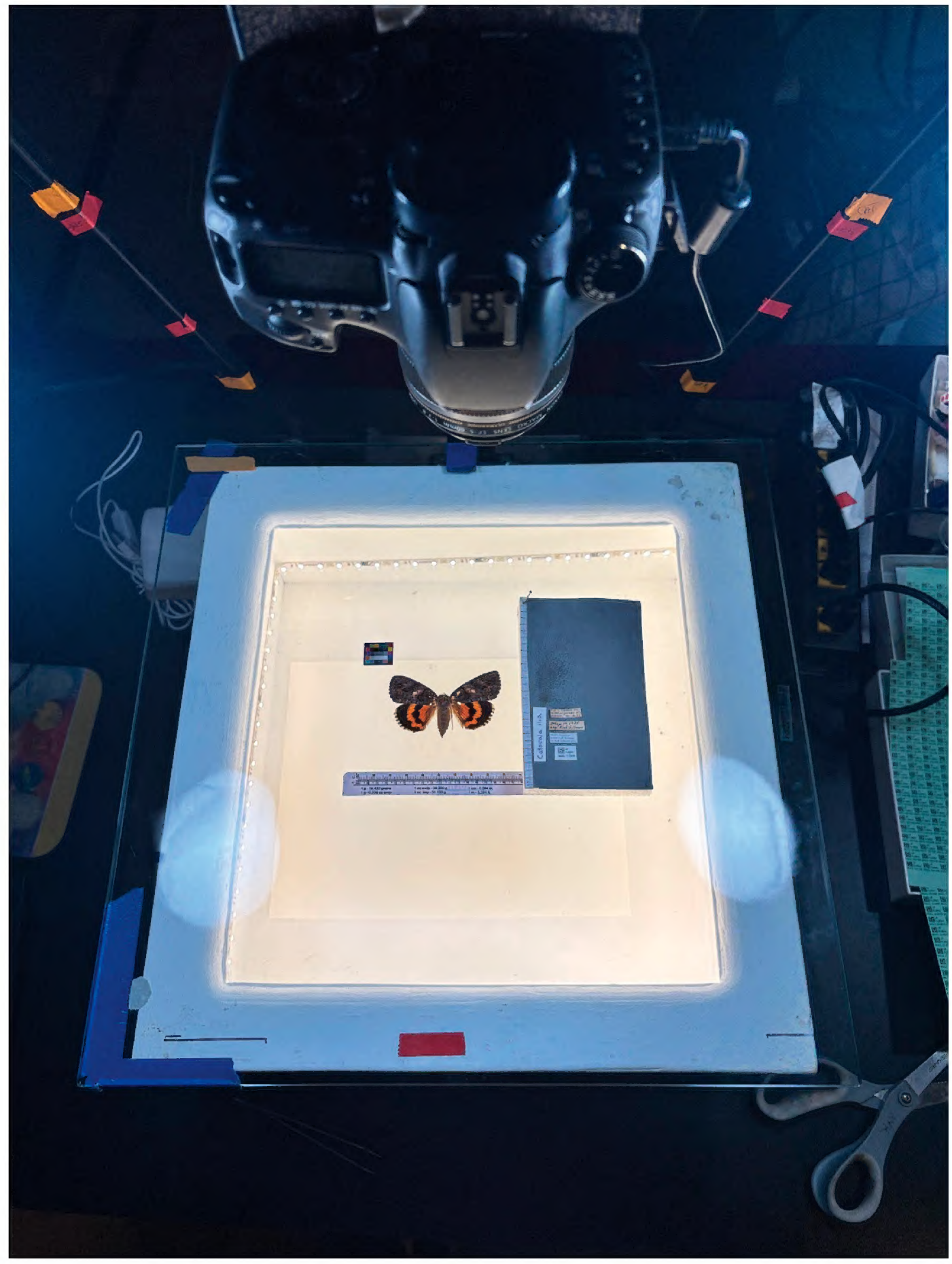

Figure 1. View of copy stand digitization from the perspective of the digitizer. The specimen was mounted in putty and placed on the tempered glass along with the color card. Label data is placed on the gray foam platform on the right along with the scientific name. The lightbox has internal LED lighting to minimize shadows. The copy stand lights are not shown but are located on both sides of the lightbox.

ningBug project (Hereld et al. 2017; Hereld and Ferrier 2019) and for capturing multispectral images of specimens (Chan et al. 2022). Similar advancements are being developed for fluid collections, enabling imaging without removing specimens from vials (Dupont et al. 2020).

Another approach to increase digitization speed is to image a whole insect drawer at once and parse single specimen images into separate records. Whole-drawer imaging usually captures an image of the specimen and associated label data. Several programs and workflows have been proposed, including InSelect (Hudson et al. 2015), GigaPan (Bertone et al. 2012), SATSCAN (Blagoderov et al. 2012; Mantle et al. 2012), and DSCAN (Schmidt et al. 2012).

Some whole-drawer imaging workflows involve taking multiple images of a drawer, using software to stitch them together, and then separating the stitched image into a unique image and identifier for each specimen (Blagoderov et al. 2012; Mantle et al. 2012; Hudson et al. 2015). Other methods create a panorama of the whole drawer from individual images (Bertone et al. 2012). However, the problem with whole drawer imaging is that it does not usually capture associated label data because the data is either blocked by the specimen above or because labels are stacked on top of each other.

After imaging and image processing, data labels need to be transcribed. Transcription is the rate limiting step in the digitization process (Guralnick et al. 2024) because of limited time and money for personnel, tediousness of parsing information into different fields, difficult to read labels, and time to double check and catch transcription errors. The two most common transcription methods are to either type directly into a database or to use computer programs to help automate transcription. One widely used platform for transcription is Notes from Nature (NfN), which enables volunteers from around the world to transcribe records and gather scientific data on the Zooniverse platform (Hill et al. 2012). Each record is transcribed multiple times and the results of the transcribers can be cleaned, reconciled, and uploaded into a database.

Despite widespread use of copy stand and transcription workflows, published or publicly accessible workflows represent only a small portion of those used by entomological collections across the world. An increasing number of NHCs in the entomological community are sharing workflows on BugFlow, a community-driven resource available on Slack and GitHub (Entomology Collection Network 2023). Publishing workflows is crucial for disseminating best practices and promoting optimization, standardization, and reproducibility within and across institutions.

The Lepidoptera of North America Network (LepNet) project, funded by the NSF, enabled 29 research collections to digitize 2.1 million specimens covering most Lepidoptera species in North America, and it created large datasets of select species to study them in depth (Seltmann et al. 2017). Here we present the workflow for pre-imaging preparation, imaging, image post-processing, and transcription used for LepNet, implemented at the McGuire Center for Lepidoptera and Biodiversity (**MGCL**), which has one of the largest collections of Lepidoptera (Kawahara et al. 2012), The workflow used a copy stand to capture dorsal and ventral images and label data of pinned-Lepidoptera specimens, and a post-processing computer scripting pipeline for renaming and uploading images and label data onto the Symbiota Collections of Arthropods Network (**SCAN**).

Materials and methods

There are four aspects that underpin our digitization protocol. 1) Image capture set up using a copy stand, lightbox, and camera; 2) Barcode label selection and application; 3) Running software and image processing scripts that streamline editing to ensure images have consistent quality and appearance; 4) Transcription of processed images. We experimented with using a copy stand to image fluid specimens and have included observations in Suppl. material 1.

Image capture setup

The image capture setup requires five key pieces of equipment: a copy stand, lightbox, foam sheet, tempered glass, and a camera. A detailed list of supplies is provided in Suppl. material 2. We used a Beseler CS-14 copy stand with a custom-built lightbox, based on the design by Ortiz-Acevedo et al. (2020), placed on the copy stand as an imaging platform (Fig. 1). The lightbox features two strips of LED lighting along the inner walls to provide backlighting and minimize shadows. Its inner base was lined with a removable neutral gray surface, made from either craft foam or a metal plate. For enhanced contrast, white or black foam backgrounds were used for specimens with gray wing margins.

Photographs were taken with a Canon EOS7D camera, with a 60-mm lens for most specimens. Larger specimens, such as *Antheraea polyphemus* (Cramer), were imaged using a 18–55 mm lens. Image quality was optimized using camera features such as white balance, aperture, shutter speed, and ISO sensitivity. Midway through the project, we incorporated the Datacolor Spyder X Elite monitor calibration software (Datacolor, Lawrenceville, New Jersey, purchased 2021) to further standardize image lighting.

Our protocol includes a color card, silicone putty, and gray foam board placed on top of tempered glass. The gray foam board was constructed by gluing a layer of gray foam over plank Styrofoam. Each component of the setup was placed in a standardized location. The color card was positioned on the glass so that it would appear in the top left corner of the image. The silicone putty used to hold the specimen was placed adjacent to the foam board, ensuring the foam board did not interfere with the specimen (Fig. 2). The data labels, barcode, ruler, and scientific name were arranged on the foam board.

Barcode label selection and application

For pinned-specimen digitization, we either used a green colored paper barcode label, measuring 19 mm x 13 mm, printed at MGCL with BarTender 10.1 (Seagull Scientific, Bellevue, WA, USA) with code 128 or a 14 mm x 10 mm plastic laminated data matrix barcode from AlphaSystems (a Hunkar Technology Company, Midlothian, VA, USA). Each label included a human- and machine-readable unique identifier, to provide essential specimen tracking and standardize image file names.

Running software and image processing scripts

The program EOS Utility version 2.11 (Canon, USA) was used to take .cr2 and .jpg images. Settings on the camera were F-stop: F/16, sensitivity to light (ISO-200), and exposure time (shutter speed 1/15 sec). For image processing, multiple custom scripts (Leopold and O'Connell 2023) were written in Python3 (Van Rossum and Drake 2009). These scripts can be used by other collections for the same tasks and/or have data matrix or code 128 barcodes. Scripts were for renaming images ("Datamatrix.py" script), reducing size of images ("Rescale.py" script), generating .csv file of images in current working directory to make a skeletal data file ("wls.py" script), and zipping images into folders for long term storage ("Zipper.py" script). The "Datamatrix.py" script (Leopold 2021) used the

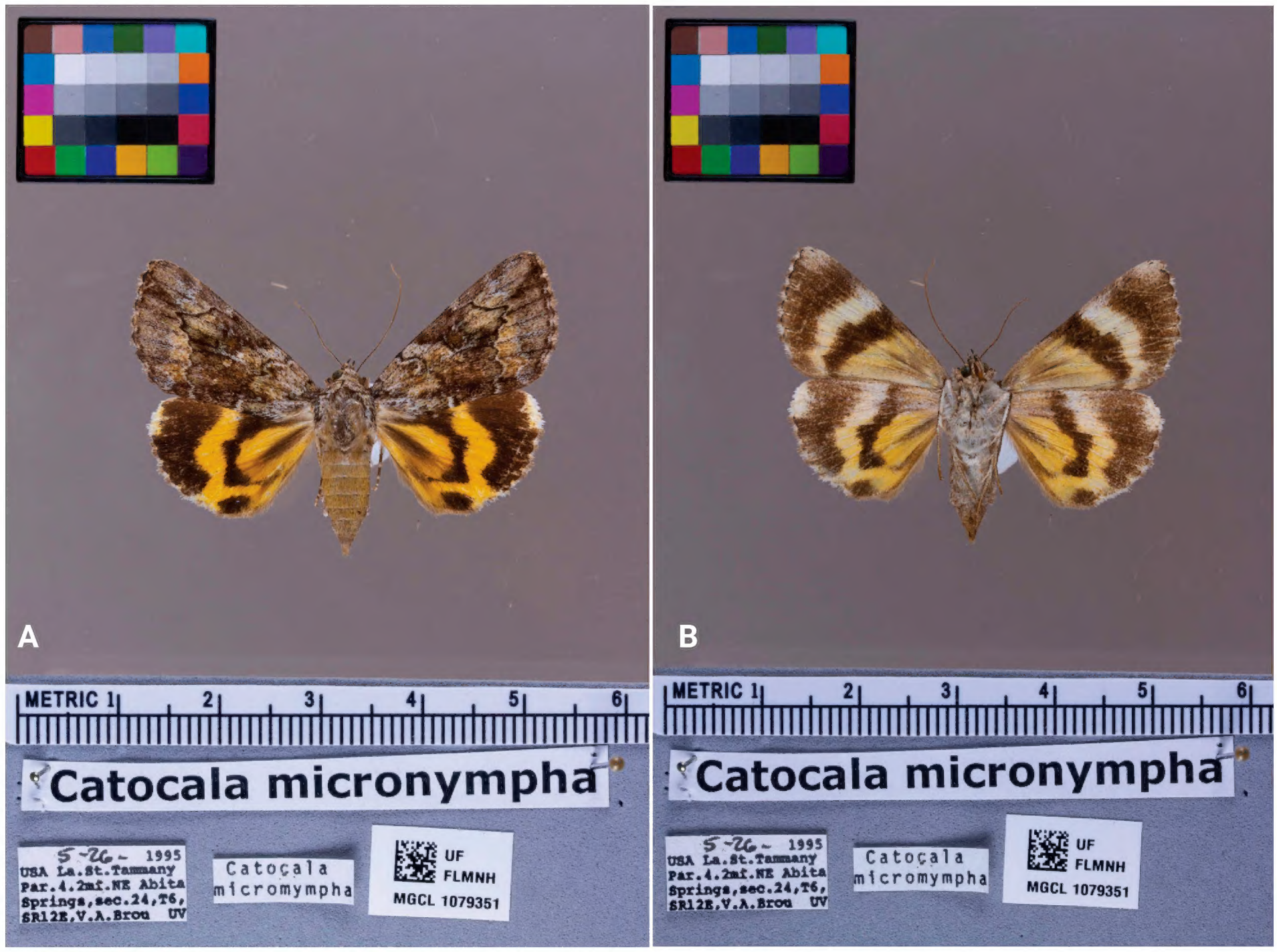

Figure 2. Example of the dorsal and ventral images of a digitized specimen (*Catocala micronympha* Guenée). Items in the picture include a color card (top left), ruler, scientific name, label data, and a unique barcode. **A.** Dorsal image; **B.** Ventral image.

libraries "dmtx" (Misbakh-Soloviov 2016) and "zbar" (Brown 2012) to extract the barcode and rename the file. The "Rescale.py" script used the PIL module, specifically the submodules "Image" and "ExifTags", to reduce the size and rotate the images (Clark 2021). The correct image orientation was determined by the value extracted from ExifTags, which was an integer between one and eight with one indicating zero degrees and eight indicating 90 degrees. The "wls.py" script was a skeletal file upload, which took the images in a file and generated a .csv file listing all file names. The scientific name was entered manually into the .csv file and uploaded onto SCAN. The "Zipper.py" script zipped specified images into a new folder of a specified size for long term archiving, sharing images with collaborators, or uploading onto SCAN.

Transcription

Records were transcribed on NfN, a digital platform aimed at engaging citizen scientists from around the world (Hill et al. 2012). Each project is called an "Expedition." Every specimen in the expedition must be transcribed by three people to be considered complete and the transcriptions were compiled into one .csv file. Transcriptions from NfN were cleaned then uploaded onto SCAN. Select records were also transcribed directly into SCAN.

Results

MGCL Digitization workflow

Our workflow consisted of four modules: 1) Pre-Imaging preparation; 2) Specimen imaging; 3) Image processing; 4) Transcription. Detailed documentation for pinned-specimen digitization is provided in Suppl. material 3. Pinned-specimen digitization workflow required 1–2 people to complete the following steps: 1. Removing the specimen from the drawer and preparing the specimen and label data for imaging. 2. Imaging. 3. Repinning data labels and data matrix barcode. 4. Returning the specimen to the drawer.

Pre-imaging preparation of pinned specimens

A project plan and specimen selection were crucial to the success of digitization. A project plan communicates to collection users all details of what specimens will be digitized, project length, and outcomes, and where specimens are temporarily stored. The MGCL is a large facility, making coordination essential to avoid disruptions to other users. Specimen selection was guided by factors such as curation quality, researcher's data needs, and institutional objectives. We moved drawers to a part of the collection near the digitization station to minimize traffic and use of the compactors. Labels with the scientific names of taxa, which were placed in the image of the corresponding species, were printed in advance.

Pinned-specimen imaging

The imaging interface, EOS Utility, was opened and a destination folder for images was created. The imaging surface of the copy stand was cleaned. The camera was adjusted to the optimal height and the camera settings, including the shutter speed, ISO, and aperture were checked.

The workflow below is described as if one person is digitizing (Fig. 3). If two people work together, one person can manage specimen related tasks and the second person can take photographs. The imaging process began with setting up the specimen. The specimen was taken out of its drawer, its labels removed from the pin, and placed on the gray foam board along with the new barcode label on the tempered glass. The specimen pin was inserted into a small piece of moldable silicone putty on the glass and the digitizer checked that the specimen's wings were level. Labels were organized on the foam board in the following manner: scientific name was placed on top with locality labels directly below and miscellaneous labels either below or to the right of locality labels. Accession and catalog number labels were placed at the very bottom. While this is a general template on how labels are arranged, the organization differed when there were very large labels associated with the specimen. In such cases, labels were arranged so that all can fit within the area allocated for photography. The dorsal side was always imaged first because the "Datamatrix.py" script appended "_D" for the first image of a barcode. The camera was then adjusted so the specimen was in focus and a dorsal image was taken simultaneously in a .cr2 and .jpg. The specimen was turned over and the head of the pin was placed in the putty, so that the ventral side was facing the camera. Labels were

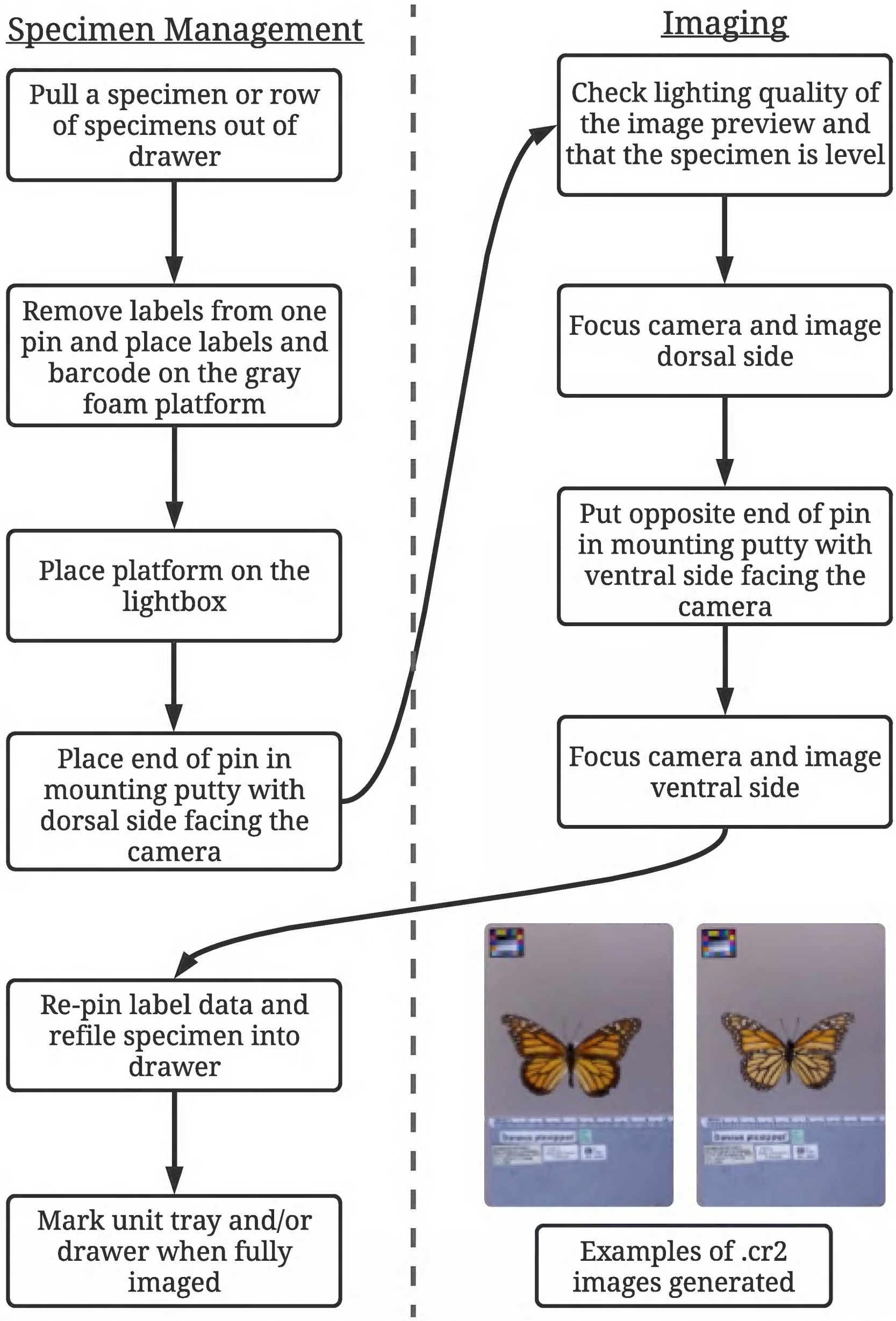
Specimen Management
Pull a specimen or row of specimens out of drawer
Remove labels from one pin and place labels and barcode on the gray foam platform
Place platform on the lightbox
Place end of pin in mounting putty with dorsal side facing the camera
Re-pin label data and refile specimen into drawer
Mark unit tray and/or drawer when fully imaged
Imaging
Check lighting quality of the image preview and that the specimen is level
Focus camera and image dorsal side
Put opposite end of pin in mounting putty with ventral side facing the camera
Focus camera and image ventral side
Examples of .cr2 images generated

Figure 3. Digitization workflow for pinned specimens. Workflow consists of specimen management and imaging tasks. Arrows trace transitions between tasks. Digitization begins with removing the specimen from the box, pulling label data from the pin, placing it onto the gray foam block, and mounting pinned specimen in the putty. The dorsal and ventral images are taken, then the label data is placed back onto the pin and the specimen is refiled. A head pin is placed to mark where imaging should begin for future sessions.

checked to confirm that information was not written or printed on the back of the labels. If information on labels was present on the backside it was imaged in the ventral photo. The camera was focused again and an image was taken. In some rare cases, a third image was taken if there were large data labels. If the specimen was pinned laterally, only one image was taken. Image capture of laterally pinned was not standardized (e.g. consistently imaging either right or left side) and is an opportunity for further standardization. Data labels were placed back on the pin. The barcode label was pinned last and faced downward. The specimen was placed back into its original drawer or unit tray. At MGCL, we created a system so that at the end of the imaging session, the digitizer placed a colored pin in the insect drawer to denote where the image session was terminated. The scientific name, number of specimens imaged, and name of the digitizer were typed in a separate Excel spreadsheet to keep track of image metadata. Imaging equipment was cleaned and stored until next use.

Pinned-specimen image processing

Four Python scripts were utilized to process images (Fig. 4). The "Datamatrix.py" script processed an image to extract a specimen catalog number from a barcode. It renamed the image file to match the barcode data and used a directional indicator that rotated the image to a standardized vertical position. The script also appended the file name to note if the image was dorsal ("_D"), ventral ("_V"), or lateral ("_L"). An image of the dorsal side was taken before the ventral side, and the script used this consistency to rename images. Lateral images were renamed by hand because the default was for the first specimen of a unique barcode to be appended with "_D". Less than one percent of specimens was renamed by hand because the barcode was unreadable. The "Rescale.py" script created a low resolution .jpg image from the original image, and the script, "Zipper.py", zipped the downscaled images into a folder. The zipped images were uploaded onto SCAN. A fourth script, "wls.py", generated a .csv file that listed all catalog numbers in each folder. The numbers were used to create a skeletal file which captured the scientific name and catalog number which was uploaded onto SCAN and linked to images. Scientific names were entered manually in the .csv file. Records with images and skeletal data on SCAN were then used for label transcription.

Pinned-specimen transcription

The main steps of an NfN expedition were specimen selection, creating a spreadsheet of data to be captured, and writing a paragraph containing information for volunteers about the expedition (Fig. 5). Expeditions had a particular theme, usually a taxonomic group to align with the needs of the institution or researcher. Once images were selected, they were downscaled using the "Rescale.py" script and zipped so that images were easier to upload to NfN. Before zipping, we made a Manifest: a .csv file that contained the catalog number and scientific name of specimens included in the NfN transcription expedition. To make the Manifest, we ran "wls.py" to capture the image file name in the folders. We manually added the scientific name and the catalog number to the .csv file, the latter of which was obtained easily from the image file name. If it was the first expedition in a series, we wrote a welcome blog, a paragraph explaining why

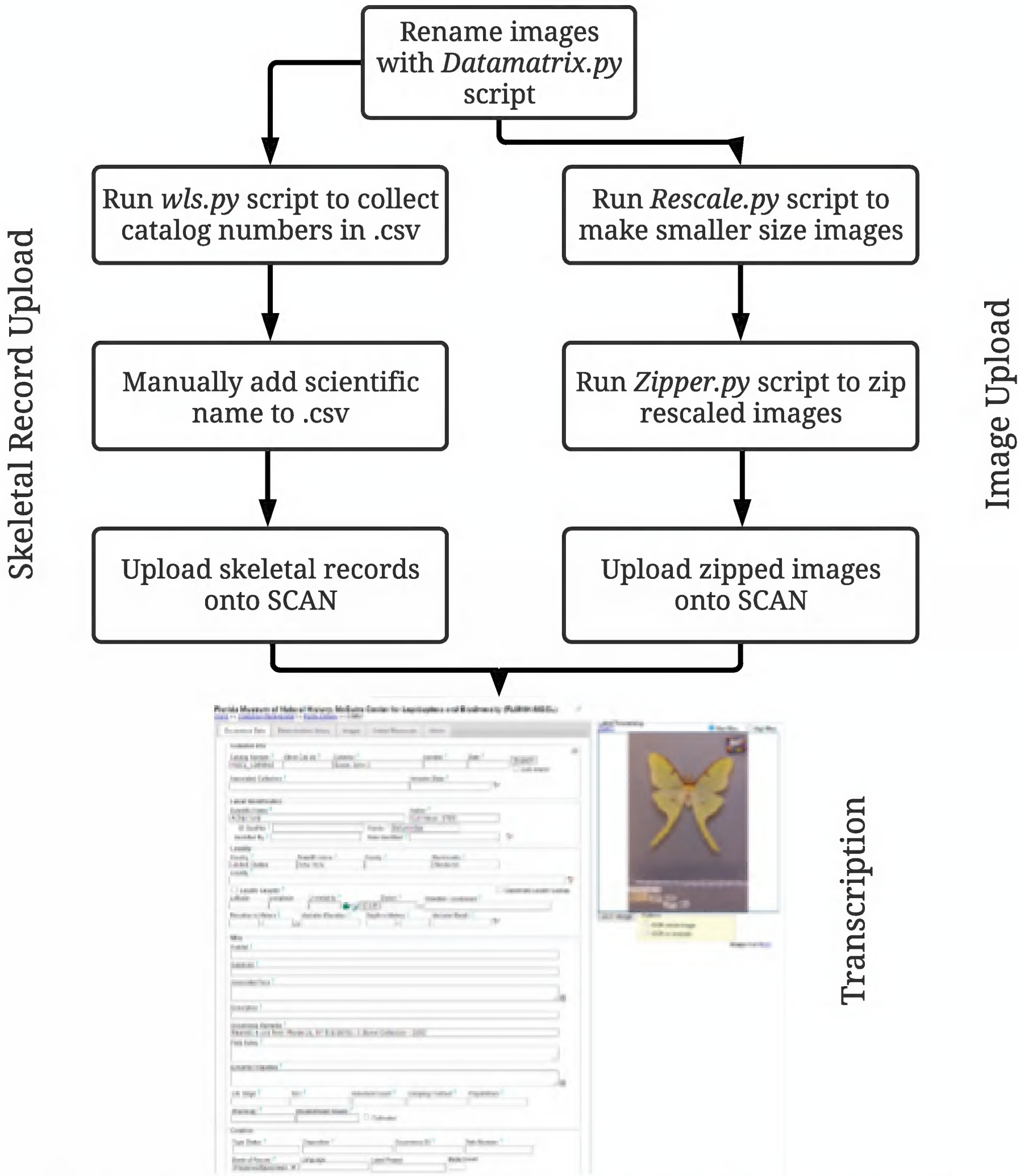
Rename images with *Datamatrix.py* script
Run *wls.py* script to collect catalog numbers in .csv
Manually add scientific name to .csv
Upload skeletal records onto SCAN
Skeletal Record Upload
Run *Rescale.py* script to make smaller size images
Run *Zipper.py* script to zip rescaled images
Upload zipped images onto SCAN
Image Upload
Transcription

Figure 4. Post-processing workflow. Image processing begins with renaming barcodes using the *Datamatrix.py* script. The renamed images are then used to generate skeletal records and resize the images using the *Rescale.py* script and zipped using *Zipper.py* then uploaded onto SCAN. Skeletal records consist of the scientific name and the catalog number. Image resolution is reduced to thumbnail quality then zipped into one folder, before being uploaded onto SCAN.

these organisms are interesting, how data will be used and some eye-catching images. We included an instruction guide for transcribers with protocols of how to parse data and transcribe in a standardized method. The Manifest, images, welcome blog, and transcription instructions were sent to NfN via email. Once the expedition was launched, we announced the expedition on the NfN blog. We engaged transcribers throughout the expedition by answering questions and engaging in conversation. Transcribers were thanked at the conclusion of the

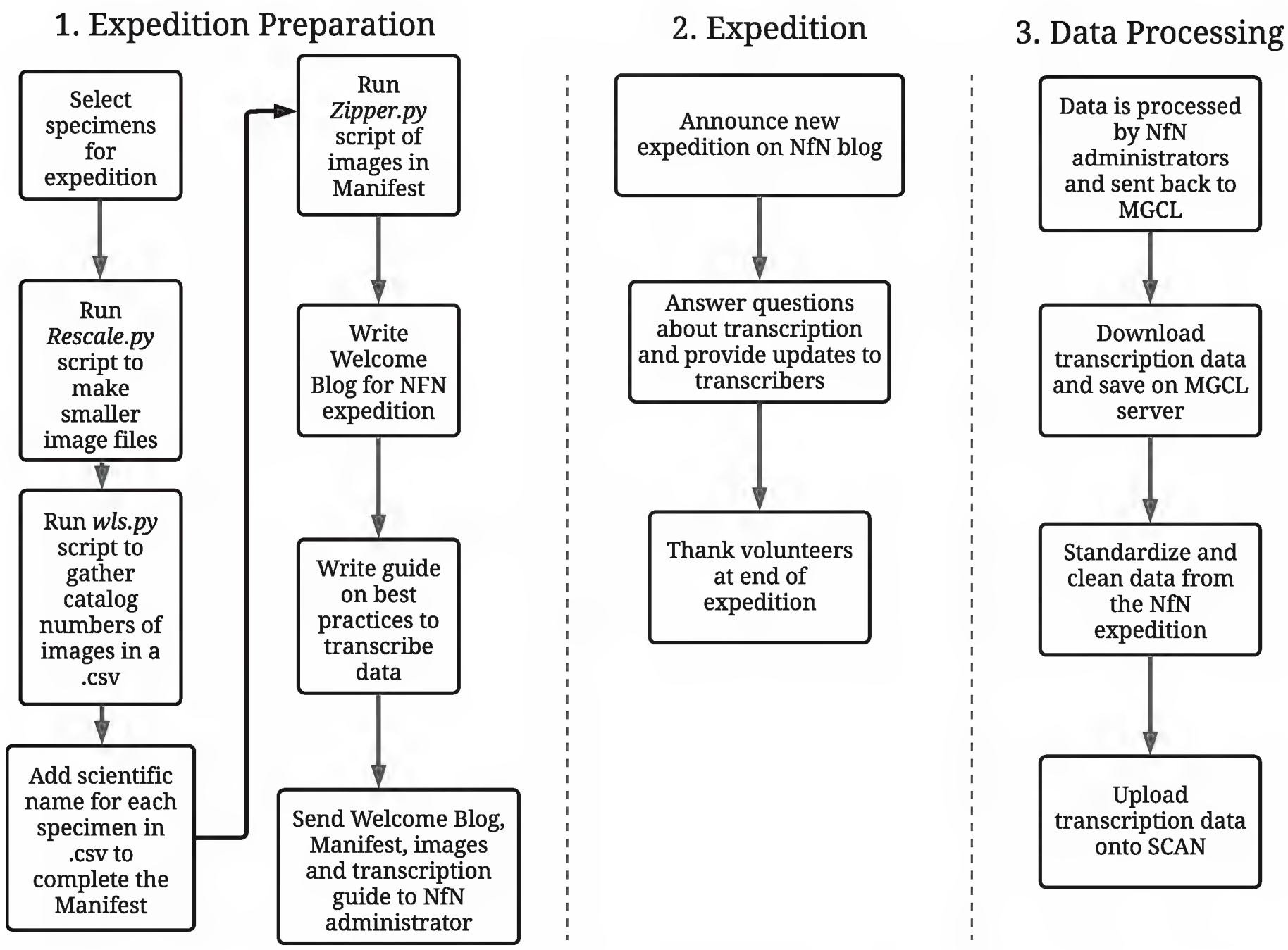
1. Expedition Preparation
Select specimens for expedition
Run *Rescale.py* script to make smaller image files
Run *wls.py* script to gather catalog numbers of images in a .csv
Add scientific name for each specimen in .csv to complete the Manifest
Run *Zipper.py* script of images in Manifest
Write Welcome Blog for NFN expedition
Write guide on best practices to transcribe data
Send Welcome Blog, Manifest, images and transcription guide to NfN administrator
2. Expedition
Announce new expedition on NfN blog
Answer questions about transcription and provide updates to transcribers
Thank volunteers at end of expedition
3. Data Processing
Data is processed by NfN administrators and sent back to MGCL
Download transcription data and save on MGCL server
Standardize and clean data from the NfN expedition
Upload transcription data onto SCAN

Figure 5. Notes from Nature workflow consists of three steps: preparing the expedition, maintenance during the expedition, and data processing. Preparation includes selecting specimens and images for transcription and writing a welcome blog and instructions on how to transcribe. During the expedition, we monitor for questions. Data processing consists of cleaning the data and formatting them so they can be uploaded into a database.

expedition. Data from all three transcriptions per record were compiled by NfN staff, then sent to MGCL. Data in the reconciled .csv file were manually cleaned column by column to standardize data then uploaded onto SCAN.

Statistics from LepNet digitization efforts

We tested our approach through the LepNet digitization efforts. Approximately 162,000 raw images of 81,000 specimens, representing ~2500 species of Lepidoptera, were digitized at the MGCL from September 2016 to May 2021. The number of images taken increased year over year from 9,872 in 2017, to 23,438 in 2018 to 37,209 in 2019. Most specimens were imaged twice (once dorsally and once ventrally). Although not a LepNet priority, on occasion, we imaged a laterally pinned specimen and its label data to enable data capture of the record. Specimens were imaged across 23 families with the majority in Erebidae (61,015 images), Saturniidae (37,301 images), Geometridae (18,903 images), and Sphingidae (17,761 images). Approximately 54,000 transcriptions (64% of LepNet specimens) were fully or partially completed. Of these 54,000 transcriptions, 21,500 were skeletal transcriptions (only genus, species, and country and/or state/province transcribed) and 32,500 were full transcriptions. The MGCL

had 23 expeditions through Notes from Nature, totaling ~18,400 specimens. Of these 18,400 specimens, 10,400 were uploaded to SCAN. The remaining 8,000 needed data cleaning and were not included in the transcription total.

From February 2019 to May 2021 (27 months), MGCL personnel involved in the project collectively worked for 70,296 minutes (1,171.6 hrs) to obtain 54,099 images (includes dorsal and ventral), an average of 23 specimens per hour (sp/hr). The rate of imaging with one digitizer ranged from 11 to 23 per hour and was ~18 sp/hr while with two digitizers, the rate ranged from 13 to 44 per hour and was ~25–26 sp/hr. Forty-one community members and/or students volunteered. There were eight staff members, four of which were students who received hands-on training in a setting that usually has few opportunities for employment. Additionally, ~30 middle- and high-school students imaged and transcribed specimens for a week-long summer camp as part of the Florida Museum of Natural History's Junior Volunteer program.

Discussion

Digitization is often defined as a series of steps from pre-imaging to georeferencing (Nelson et al. 2012). Digitization requires decision making to optimize data-capture rate while maintaining desired data quality standards. Decisions related to the rationales and tradeoffs in a workflow are seldom discussed in papers. Our workflow was developed through extensive experimentation to determine the most effective equipment and settings for digitization and to evaluate their impact on reproducibility. This process involved iterative adjustments to equipment configurations and setup variables. Here, we discuss the benefits and drawbacks of the pinned-digitization workflow.

Lightbox and copy stand

The combination of a lightbox and a copy stand for digitization creates a user-friendly and flexible method to image Lepidoptera specimens. The setup can accommodate most butterflies and large moths; increased magnification is required for imaging micromoths. Our digitization set up can image large specimens and specimens with unique morphologies such as those with long tails on their hindwings (e.g., some Saturniidae) by imaging on the glass surface or in the interior of the lightbox. This is due to the fact that the height of the camera mount on the copy stand can be adjusted. A lightbox and copy stand can be set up at minimal cost and can be used to generate images quickly. Once set up, the equipment is easy to maintain and use. The benefit of designing a customized lightbox to use in-house is that it can be made to desired optimal specifications, usually at a lower cost point. Lightboxes at MGCL were made of readily available supplies, including a wood frame coated in white paint, with LED strips glued and soldered on the inside for illuminating specimens during digitization. The main drawbacks with a copy stand in the MGCL digitization workflow centered around lighting. It is preferable to have all digitization stations in a single workspace, but in practice, it is often necessary to have digitization stations in different workspaces that inherently have variable lighting conditions. There are multiple sources of lighting to control, including copy stand lightbulbs, LED strips, and ceiling lighting. The MGCL had three digitiza-

tion stations, each in a different room and with different ambient lighting, which required additional methods to ensure the accuracy of specimen coloration. For example, wax paper was put over the bulb fixtures in order to remove excessive glare of the light from the copy stand fixtures. Light- and color-calibration was performed with DataColor Spyder X Elite for each digitization computer monitor to standardize lighting between imaging stations. Another issue was that specimens of different sizes required different lighting angles and camera settings. Digitizers kept track of camera settings and the angles and heights of lights to standardize image capture throughout each project.

Pinned-specimen imaging

The goal of the pinned-specimen imaging workflow was to capture specimen images and label data. For spread specimens, dorsal and ventral images were taken. For field-pinned specimens with unspread wings, a single lateral image was taken. We decided to image the dorsal and ventral sides of specimens because the ventral side of lepidopterans often has species-specific characteristics such as eye spots and other markings that are absent on the dorsal side. The main workflow decisions centered around specimen handling, data gathering, and speed and imaging efficiency.

We prioritized minimizing contact with specimens to ensure their long-term their preservation by handling specimens only once. To maximize the information obtained, we captured dorsal and ventral images along with associated label data. Although imaging both sides of specimens required more time initially, our approach reduced overall handling and proved more efficient in the long term.

We also had to decide whether to image label data separately or simultaneously with the specimen. Ultimately, we chose to capture label data simultaneously in images in both dorsal and ventral images. This approach was based on the idea that the images should aid in identification and analyses, rather than serve as publication-quality outputs, while also facilitating transcription. Another benefit of capturing the label data and specimen in the same image is that the photo and transcription can be checked against each other by future researchers. However, a drawback of this method is that the specimen occupies a smaller percentage of the image due to the inclusion of label data. One advantage of our setup is that information on data labels can be easily captured and does not require additional images to be taken. By standardizing the placement of label data, we aim to simplify the re-pinning process and reduce the need for searching for the desired label during transcription.

Data labels were placed on a platform ~0.75 inches tall, constructed from plank Styrofoam. This material was chosen for its durability – it resists breaking or denting while remaining malleable enough to accommodate pins. The elevated platform ensures that the label data and the barcode are in the same focal plane as the specimen.

Another consideration was whether imaging was more efficient solo or in tandem. A single digitizer averaged 18 sp/hr, which could be as high as 23 sp/hr if the digitizer was highly experienced. In contrast, two-person teams averaged 23–24 to upwards of 44 specimens per hour. The average difference in efficiency is minimal because, in tandem setups, one person may work faster than the other, leading to brief idle periods. However if two experienced imag-

ers are working together the average could be significantly greater than normal. Tandem imaging offers benefits such as task specialization, allowing each person to focus on fewer responsibilities, and reducing feelings of isolation—an important factor for volunteers. Additionally, imaging efficiency varies with specimen size; smaller specimens are quicker to image than large specimens.

Pinned-specimen digitization image processing

Images were processed every few weeks, with editing taking a few hours, primarily due to the time required for scripts to run. The "Datamatrix.py" script failed to rename some barcodes because there was debris on the barcode, a pinhole through the barcode, or part of the machine-readable barcode being either out of focus or missing from the image. In these instances, image files were renamed manually. For large Lepidoptera, such as *Attacus* L. spp. or *Antheraea polyphemus*, the camera had to be mounted higher, resulting in barcodes appearing too small for successful automated renaming.

Pinned-specimen digitization transcription

Transcription in NfN had its advantages and disadvantages, with the primary considerations being time and quality. The NfN is a more time-intensive process for the expedition coordinator, requiring significant effort to prepare images for the expedition and to clean and upload the data afterward. Smaller projects may not justify the time investment, while larger projects risk prolonged completion times.

Transcriptions from NfN exhibited greater variability, requiring significant effort to standardize. This was because transcribers often entered label data exactly as it appeared, with fields like locality frequently abbreviated in inconsistent ways. To streamline formatting, we applied batch changes in Excel to standardize fields such as names, dates, and localities. For efficiency, we prioritized cleaning frequently repeated data, such as locality and collector names. Additionally, we added scientific names before the expedition to ensure accuracy and minimize transcription error. The focus on our project was developing the imaging infrastructure, and there are opportunities to improve the transcription workflow and better understand the efficacy of transcriptions from NfN. A future direction is to incorporate technologies and software to expedite cleaning. Options include OpenRefine software to clean and reconcile messy data (Delpeuch et al. 2024) or to incorporate Label Babel, recently developed machine learning optical character recognition methods integrated with NfN (Guralnick et al. 2024). Another emerging approach is to use artificial intelligence and large language models, to do the initial transcription then check these data by humans (Herbst et al. 2025).

Fluid digitization on a copy stand

We conducted preliminary tests to assess whether a copy stand could be used to image an individual fluid specimen. In our tests, a specimen and data labels were removed from the jar and partially dried, then imaged on the copy stand (Suppl. material 1). We believe that this method may be useful to image individual specimens but requires further testing to optimize image quality and scalability of the method.

Conclusion

We provide a workflow for digitization of Lepidoptera that covers pre-imaging preparation, imaging, image processing, and transcription. Our workflow utilizes a copy stand for pre-imaging digitization tasks and imaging pinned specimens and includes image processing and transcription steps, the latter two which utilize bioinformatic scripts for optimization. Our approach is fast and inexpensive compared to other existing digitization workflows, and it has been used in research projects, such as a study on *Catocala* Schrank wing color and polymorphism (Homziak 2022). Many entomology collections utilize copy stand workflows that have the same overarching steps but a myriad of ways to accomplish them. It is critical for global digitization success and efficiency to share open access workflows through publication, institutions, and community-driven resources. We hope that the sharing of workflows will further improve digitization best practices.

Acknowledgements

We thank Kevin Keegan and two anonymous reviewers, whose comments greatly enhanced the manuscript. This project would not have been possible without help from many people at the McGuire Center. We thank Caroline Storer, who provided critical feedback on early drafts. Violet Butterwort, Jaret Daniels, Emily Ellis, James Hayden, Lillian Hendrick, Amanda Markee, Jacqueline Miller, Andrei Sourakov, and Keith Willmott provided valuable insight on the digitization workflow and curation of specimens. We also thank Ana Paula dos Santos de Carvalho, Nicholas Homziak, Jose Martinez, and Ryan St Laurent for curation of taxa prior to digitization and feedback on the workflow. Digitization would not have been possible without the help of our many volunteers, but we would like to especially thank Jim Grantham for his many years of volunteer service.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

No use of AI was reported.

Funding

LepNet was funded by NSF grant DBI 1601369 to AYK.

Author contributions

Conceptualization: SLH. Data curation: SLH, LK. Funding acquisition: AYK. Methodology: TD, NNVT, AP, EJH, LK, SLH. Project administration: LK, SLH. Software: AL. Supervision: LK, SLH. Order for Writing review and editing: LK, DP, AYK, AL, EJH, NNVT, TD, AP. Writing – original draft: LK. Writing – review and editing: DP.

Author ORCIDs

Laurel Kaminsky https://orcid.org/0000-0002-0786-8350
Ngoc-Nhu V. Tran https://orcid.org/0000-0002-2555-2958
Anupama Priyadarshini https://orcid.org/0009-0003-2009-5940
David Plotkin https://orcid.org/0000-0002-2339-655X
Akito Y. Kawahara https://orcid.org/0000-0002-3724-4610

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Aiello BR, Sikandar UB, Minoguchi H, Bhinderwala B, Hamilton CA, Kawahara AY, Sponberg S (2021) The evolution of two distinct strategies of moth flight. Journal of the Royal Society, Interface 18(185): e20210632. https://doi.org/10.1098/rsif.2021.0632

Almryad AS, Kutucu H (2020) Automatic identification for field butterflies by convolutional neural networks. Engineering Science and Technology, an International Journal 23(1): 189–195. https://doi.org/10.1016/j.jestch.2020.01.006

Ball-Damerow JE, Brenskelle L, Barve N, Soltis PS, Sierwald P, Bieler R, LaFrance R, Ariño AH, Guralnick RP (2019) Research applications of primary biodiversity databases in the digital age. PLoS ONE 14(9): e0215794. https://doi.org/10.1371/journal.pone.0215794

Bedford FE, Whittaker RJ, Kerr JT (2012) Systemic range shift lags among a pollinator species assemblage following rapid climate change. Botany 90(7): 587–597. https://doi.org/10.1139/b2012-052

Belitz MW, Larsen EA, Shirey V, Li D, Guralnick RP (2023) Phenological research based on natural history collections: Practical guidelines and a lepidopteran case study. Functional Ecology 37(2): 234–247. https://doi.org/10.1111/1365-2435.14173

Bertone MA, Blinn RL, Stanfield TM, Dew KJ, Seltmann KC, Deans AR (2012) Results and insights from the NCSU Insect Museum GigaPan project. ZooKeys 209: 115–132. https://doi.org/10.3897/zookeys.209.3083

Blagoderov V, Kitching IJ, Livermore L, Simonsen TJ, Smith VS (2012) No specimen left behind: Industrial scale digitization of natural history collections. ZooKeys 209: 133–146. https://doi.org/10.3897/zookeys.209.3178

Blagoderov V, Penn M, Sadka M, Hine A, Brooks S, Siebert DJ, Sleep C, Cafferty S, Cane E, Martin G, Toloni F, Wing P, Chainey J, Duffell L, Huxley R, Ledger S, McLaughlin C, Mazzetta G, Perera J, Crowther R, Douglas L, Durant J, Honey M, Huertas B, Howard T, Carter V, Albuquerque S, Paterson G, Kitching IJ (2017) iCollections methodology: Workflow, results and lessons learned. Biodiversity Data Journal 19893: e19893. https://doi.org/10.3897/BDJ.5.e19893

Brown J (2012) ZBar. https://github.com/ZBar/ZBar [accessed 12 June 2019]

Buschbacher K, Ahrens D, Espeland M, Steinhage V (2020) Image-based species identification of wild bees using convolutional neural networks. Ecological Informatics 55: e101017. https://doi.org/10.1016/j.ecoinf.2019.101017

Caspers M, Willemse L, Miracle EG, van Nieukerken EJ (2019) Butterflies in bags: Permanent storage of Lepidoptera in glassine envelopes. Nota Lepidopterologica 42(1): 1–16. https://doi.org/10.3897/nl.42.28654

Chan W-P, Childers RR, Ashe S, Tsai C-C, Elson C, Keleher KJ, Sipe RLH, Maier CA, Sourakov A, Gall LF, Bernard GD, Soucy ER, Yu N, Pierce NE (2022) A high-throughput mul-

tispectral imaging system for museum specimens. Communications Biology 5(1): e1318. https://doi.org/10.1038/s42003-022-04282-z

Childers RR, Cai L, Chowdhury S, Crall J, Cornwall M, Sipe RH, Maier C, Maunsell S, Talavera G, Tsai C-C, Angier K, Barve V, Dankowicz E, Hinolan J, Itliong M, Naive MA, Johnson G, Matos F, Morgan V, Plotkin D, Salcedo M, Shirey V, Vermilya K, Winstanley J, Wu A, Guralnick R, Kawahara A, Lohman D, Ries L, Soucy E, Vila R, Yu N, Pierce N, Chan W-P (2023) Selection on size has generated distinctive paired wing flight systems for butterfly flight and migration. https://doi.org/10.21203/rs.3.rs-2968821/v1

Clark JA (2021) Pillow (PIL Fork) 8.2.0. https://pypi.org/project/Pillow/8.2.0/ [accessed 12 June 2019]

Cobb NS, Gall LF, Zaspel JM, Dowdy NJ, McCabe LM, Kawahara AY (2019) Assessment of North American arthropod collections: Prospects and challenges for addressing biodiversity research. PeerJ 7: e8086. https://doi.org/10.7717/peerj.8086

Cunha C, Narotamo H, Monteiro A, Silveira M (2023) Detection and measurement of butterfly eyespot and spot patterns using convolutional neural networks. PLoS ONE 18(2): e0280998. https://doi.org/10.1371/journal.pone.0280998

Delpeuch A, Morris T, Huynh D, Weblate, Mazzocchi S, Jacky, Guidry T, elebitzero, Stephens O, Matsunami I, Sproat I, Larsson A, Santos S, allanaaa, kushthedude, Fauconnier S, Mishra E, Magdinier M, Beaubien A, Liu L, Giroud F, Ong J, Tacchelli F, Nordhøy A, Luca Martinelli [Sannita], Kanye E, Saby M, Chandra L (2024) OpenRefine/OpenRefine: OpenRefine v3.8-beta1. https://zenodo.org/records/10689569 [accessed 20 October 2025]

DeWalt R, Yoder M, Snyder E, Dmitriev D, Ower G (2018) Wet collections accession: A workflow based on a large stonefly (Insecta, Plecoptera) donation. Biodiversity Data Journal 6: e30256. https://doi.org/10.3897/BDJ.6.e30256

Dupont S, Humphries J, Butcher A, Baker E, Balcells L, Price B (2020) Ahead of the curve: Three approaches to mass digitisation of vials with a focus on label data capture. Research Ideas and Outcomes 6: e53606. https://doi.org/10.3897/rio.6.e53606

Earl C, Belitz MW, Laffan SW, Barve V, Barve N, Soltis DE, Allen JM, Soltis PS, Mishler BD, Kawahara AY, Guralnick R (2021) Spatial phylogenetics of butterflies in relation to environmental drivers and angiosperm diversity across North America. iScience 24(4): e102239. https://doi.org/10.1016/j.isci.2021.102239

Echevarría Ramos M, Hulshof CM (2019) Using digitized museum collections to understand the effects of habitat on wing coloration in the Puerto Rican monarch. Biotropica 51(4): 477–483. https://doi.org/10.1111/btp.12680

Entomology Collection Network (2023) BugFlow. https://github.com/EntCollNet/BugFlow [accessed 13 November 2023]

GBIF (2024) The Global Biodiversity Information Facility What is GBIF? https://www.gbif.org/what-is-gbif [accessed 20 September 2024]

Giribet G, Edgecombe GD (2012) Reevaluating the arthropod tree of life. Annual Review of Entomology 57(1): 167–186. https://doi.org/10.1146/annurev-ento-120710-100659

Guralnick R, LaFrance R, Denslow M, Blickhan S, Bouslog M, Miller S, Yost J, Best J, Paul DL, Ellwood E, Gilbert E, Allen J (2024) Humans in the loop: Community science and machine learning synergies for overcoming herbarium digitization bottlenecks. Applications in Plant Sciences 12(1): e11560. https://doi.org/10.1002/aps3.11560

Hamilton CA, Winiger N, Rubin JJ, Breinholt J, Rougerie R, Kitching IJ, Barber JR, Kawahara AY (2022) Hidden phylogenomic signal helps elucidate Arsenurine silkmoth phylogeny and the evolution of body size and wing shape trade-offs. Systematic Biology 71(4): 859–874. https://doi.org/10.1093/sysbio/syab090

Harris KM, Marsico TD (2017) Digitizing specimens in a small herbarium: A viable workflow for collections working with limited resources. Applications in Plant Sciences 5(4): e1600125. https://doi.org/10.3732/apps.1600125

Hedrick BP, Heberling JM, Meineke EK, Turner KG, Grassa CJ, Park DS, Kennedy J, Clarke JA, Cook JA, Blackburn DC, Edwards SV, Davis CC (2020) Digitization and the future of natural history collections. Bioscience 70(3): 243–251. https://doi.org/10.1093/biosci/biz163

Herbst R, Stille D, Gwilliam GF, Von Konrat M, Campbell T, Gaswick W, Grewe F, Hansen K, Iacobelli F, Jellema K, Kawasaki ML, Kreider D, Ree RH, Rodriguez Y, Wolpert A (2025) Unlocking the Past: The Potential of Large Language Models to Revolutionize Transcription of Natural History Collections. Data Intelligence 7(2): 237–264. https://doi.org/10.3724/2096-7004.di.2025.0011

Hereld M, Ferrier N (2019) LightningBug ONE: An experiment in high-throughput digitization of pinned insects. Biodiversity Information Science and Standards 3: e37228. https://doi.org/10.3897/biss.3.37228

Hereld M, Ferrier NJ, Agarwal N, Sierwald P (2017) Designing a High-Throughput Pipeline for Digitizing Pinned Insects. 2017 IEEE 13th International Conference on e-Science (e-Science). IEEE, Auckland, 542–550. https://doi.org/10.1109/eScience.2017.88

Hill A, Guralnick R, Smith A, Sallans A, Gillespie R, Denslow M, Gross J, Murrell Z, Conyers T, Oboyski P, Ball J, Thomer A, Prys-Jones R, De La Torre J, Kociolek P, Fortson L (2012) The notes from nature tool for unlocking biodiversity records from museum records through citizen science. ZooKeys 209: 219–233. https://doi.org/10.3897/zookeys.209.3472

Holmes MW, Hammond TT, Wogan GOU, Walsh RE, LaBarbera K, Wommack EA, Martins FM, Crawford JC, Mack KL, Bloch LM, Nachman MW (2016) Natural history collections as windows on evolutionary processes. Molecular Ecology 25(4): 864–881. https://doi.org/10.1111/mec.13529

Homziak NT (2022) Evolution of hindwing color in underwing moths. PhD Thesis, University of Florida, Gainesville, Florida, United States.

Hudson LN, Blagoderov V, Heaton A, Holtzhausen P, Livermore L, Price BW, van der Walt S, Smith VS (2015) Inselect: Automating the digitization of natural history collections. PLoS ONE 10(11): e0143402. https://doi.org/10.1371/journal.pone.0143402

iDigBio.org (2024) iDigBio Home Page. https://www.idigbio.org/ [accessed 20 September 2024]

Kawahara AY, Emmel TC, Miller J, Warren AD (2012) A new institution devoted to insect science: The Florida Museum of Natural History, McGuire Center for Lepidoptera and Biodiversity. Insect Science 19(3): 426–428. https://doi.org/10.1111/j.1744-7917.2011.01490.x

Kharouba HM, Lewthwaite JMM, Guralnick R, Kerr JT, Vellend M (2019) Using insect natural history collections to study global change impacts: Challenges and opportunities. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 374(1763): e20170405. https://doi.org/10.1098/rstb.2017.0405

Leopold A (2021) FLMNH-MGCL/datamatrix: v0.1.0. https://doi.org/10.5281/zenodo.4829080 [accessed 15 February 2025]

Leopold A, O'Connell M (2023) FLMNH-MGCL/digitization: 0.0.1. https://zenodo.org/records/10032175 [accessed 15 February 2025]

Li D, Stucky BJ, Deck J, Baiser B, Guralnick RP (2019) The effect of urbanization on plant phenology depends on regional temperature. Nature Ecology & Evolution 3(12): 1661–1667. https://doi.org/10.1038/s41559-019-1004-1

Mantle B, LaSalle J, Fisher N (2012) Whole-drawer imaging for digital management and curation of a large entomological collection. ZooKeys 209: 147–163. https://doi.org/10.3897/zookeys.209.3169

May KJ, Ristaino JB (2004) Identity of the mtDNA haplotype(s) of *Phytophthora infestans* in historical specimens from the Irish Potato Famine. Mycological Research 108(5): 471–479. https://doi.org/10.1017/S0953756204009876

Mendez PK, Lee S, Venter CE (2018) Imaging natural history museum collections from the bottom up: 3D print technology facilitates imaging of fluid-stored arthropods with flatbed scanners. ZooKeys 795: 49–65. https://doi.org/10.3897/zookeys.795.28416

Misbakh-Soloviov VA (2016) dmtx/dmtx-utils. https://github.com/dmtx/dmtx-utils [accessed 11 June 2019]

Nelson G, Ellis S (2018) The history and impact of digitization and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B 374(1763): e20170391. https://doi.org/10.1098/rstb.2017.0391

Nelson G, Paul D, Riccardi G, Mast AR (2012) Five task clusters that enable efficient and effective digitization of biological collections. ZooKeys 209: 19–45. https://doi.org/10.3897/zookeys.209.3135

Nelson G, Sweeney P, Wallace LE, Rabeler RK, Allard D, Brown H, Carter JR, Denslow MW, Ellwood ER, Germain-Aubrey CC, Gilbert E, Gillespie E, Goertzen LR, Legler B, Marchant DB, Marsico TD, Morris AB, Murrell Z, Nazaire M, Neefus C, Oberreiter S, Paul D, Ruhfel BR, Sasek T, Shaw J, Soltis PS, Watson K, Weeks A, Mast AR (2015) Digitization workflows for flat sheets and packets of plants, algae, and fungi. Applications in Plant Sciences 3(9): e1500065. https://doi.org/10.3732/apps.1500065

Ortiz-Acevedo E, Gomez JP, Espeland M, Touissaint EFA, Willmott KR (2020) The roles of wing color pattern and geography in the evolution of Neotropical Preponini butterflies. Ecology and Evolution 23(23): 12801–12816. https://doi.org/10.1002/ece3.6816

Owens HL, Lewis DS, Condamine FL, Kawahara AY, Guralnick RP (2020) Comparative Phylogenetics of *Papilio* butterfly wing shape and size demonstrates independent hindwing and forewing evolution. Systematic Biology 69(5): 813–819. https://doi.org/10.1093/sysbio/syaa029

Paterson G, Albuquerque S, Blagoderov V, Brooks S, Cafferty S, Cane E, Carter V, Chainey J, Crowther R, Douglas L, Durant J, Duffell L, Hine A, Honey M, Huertas B, Howard T, Huxley R, Kitching I, Ledger S, McLaughlin C, Martin G, Mazzetta G, Penn M, Perera J, Sadka M, Scialabba E, Self A, Siebert DJ, Sleep C, Toloni F, Wing P (2016) iCollections – Digitising the British and Irish Butterflies in the Natural History Museum, London. Biodiversity Data Journal 9559: e9559. https://doi.org/10.3897/BDJ.4.e9559

Popkov A, Konstantinov F, Neimorovets V, Solodovnikov A (2022) Machine learning for expert-level image-based identification of very similar species in the hyperdiverse plant bug family Miridae (Hemiptera: Heteroptera). Systematic Entomology 47(3): 487–503. https://doi.org/10.1111/syen.12543

Qin W, Abbas A, Abbas S, Alam A, Chen D, Hafeez F, Ali J, Romano D, Chen R-Z (2024) Automated lepidopteran pest developmental stages classification via transfer learning framework. Environmental Entomology 53(6): 1062–1077. https://doi.org/10.1093/ee/nvae085

Schmidt S, Balke M, Lafogler S (2012) DScan – a high-performance digital scanning system for entomological collections. ZooKeys 209: 183–191. https://doi.org/10.3897/zookeys.209.3115

Seltmann KC, Cobb NS, Gall LF, Bartlett CR, Basham MA, Betancourt I, Bills C, Brandt B, Brown RL, Bundy C, Caterino MS, Chapman C, Cognato A, Colby J, Cook SP, Daly KM,

Dyer LA, Franz NM, Gelhaus JK, Grinter CC, Harp CE, Hawkins RL, Heydon SL, Hill GM, Huber S, Johnson N, Kawahara AY, Kimsey LS, Kondratieff BC, Krell F-T, Leblanc L, Lee S, Marshall CJ, Mccabe LM, Mchugh JV, Menard KL, Opler PA, Palffy-Muhoray N, Pardikes N, Peterson MA, Pierce NE, Poremski A, Sikes DS, Weintraub JD, Wikle D, Zaspel JM, Zolnerowich G (2017) LepNet: The Lepidoptera of North America Network. Zootaxa 4247(1): 73–77. https://doi.org/10.11646/zootaxa.4247.1.10

Short AEZ, Dikow T, Moreau CS (2018) Entomological collections in the age of big data. Annual Review of Entomology 63(1): 513–530. https://doi.org/10.1146/annurev-ento-031616-035536

Takano A, Horiuchi Y, Fujimoto Y, Aoki K, Mitsuhashi H, Takahashi A (2019) Simple but long-lasting: A specimen imaging method applicable for small- and medium-sized herbaria. PhytoKeys 118: 1–14. https://doi.org/10.3897/phytokeys.118.29434

Tegelberg R, Mononen T, Saarenmaa H (2014) High-performance digitization of natural history collections: Automated imaging lines for herbarium and insect specimens. Taxon 63(6): 1307–1313. https://doi.org/10.12705/636.13

Tegelberg R, Kahanpaa J, Karppinen J, Mononen T, Wu Z, Saarenmaa H (2017) Mass Digitization of Individual Pinned Insects Using Conveyor-Driven Imaging. 2017 IEEE 13th International Conference on e-Science (e-Science). IEEE, Auckland, 523–527. https://doi.org/10.1109/eScience.2017.85

Terry CN, Alonso-Rodríguez AM, Miller SE, Hulshof CM (2023) Lepidoptera research in Puerto Rico: Reconnecting with historical legacies to guide future priorities. Biotropica 55(6): 1215–1232. https://doi.org/10.1111/btp.13278

Thiers BM, Tulig MC, Watson KA (2016) Digitization of The New York Botanical Garden Herbarium. Brittonia 68(3): 324–333. https://doi.org/10.1007/s12228-016-9423-7

Van Rossum G, Drake FL (2009) Python 3 Reference Manual. CreateSpace, Scotts Valley, California, 242 pp.

Wells CN, Tonkyn DW (2014) Range collapse in the Diana fritillary, *Speyeria diana* (Nymphalidae). Insect Conservation and Diversity 7(4): 365–380. https://doi.org/10.1111/icad.12059

Supplementary material 1

Notes on fluid digitization with a copy stand

Authors: Laurel Kaminsky, Erin Jane Howard, Ngoc-Nhu V. Tran

Data type: docx

Explanation note: Observations of preliminary tests on how individual specimens removed from fluid can be digitized using a copy stand. Contains a workflow, at a conceptual level, for the process of specimen removal, imaging, and optimization of lighting and drying a specimen to minimize glare.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1264.134756.suppl1

Supplementary material 2

Supplies for pinned-specimen digitization

Authors: Laurel Kaminsky, Stacey Huber

Data type: docx

Explanation note: List of supplies used in the copy stand digitization of pinned Lepidoptera for the McGuire Center for Lepidoptera and Biodiversity.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1264.134756.suppl2

Supplementary material 3

MGCL pinned-specimen digitization workflow

Authors: Laurel Kaminsky, Stacey Huber, Trudi Deuel, Ngoc-Nhu V. Tran, Erin Jane Howard, Aaron Leopold, Anupama Priyadarshini

Data type: docx

Explanation note: Digitization workflow used to image pinned specimens. Contains in depth details on how to set up for imaging, image processing, data management, and upload onto SCAN. This document is aimed for the digitization manager and troubleshooting problems.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1264.134756.suppl3